

# *SAMGrid Overview*

**Adam Lyon**  
(FNAL/CD)

**CDF Grid Workshop**  
**10 March 2005**

# *SAMGrid Management*

---

- ◆ Co-project managers Wyatt Merritt and Rick St. Denis have stepped down.
  - ❖ Wyatt has stepped down as Run II Data Handling group leader
- ◆ I am the new SAMGrid project leader (as of two weeks ago)
  - ❖ Define project goals, set tasks, track work, interact with customers and stakeholders, interact with line managers (more than just Run II DH)
- ◆ Krzysztof Genser is the new Run II DH group leader
  - ❖ Manages the people (all of whom are on the SAMGrid team), is a member of the SAMGrid team, leading CDF SAM deployment

# *Management changes*

---

- ◆ **My science is DØ, but my SAMGrid is CD**
  - ❖ Aside from being a DØ sam shifter, I have no DH responsibilities at DØ (nor anywhere else)
- ◆ Improve planning, documentation (both internal and external), and process
- ◆ Improve team cohesiveness
- ◆ Institute a more formal relationship with customers/stakeholders
- ◆ It's taking me awhile to ramp up...
  - ❖ What follows are my thoughts and ideas. Further reflection and alterations are likely...

# *SAMGrid Requirements*

---

- ◆ Transfer enormous amounts of data needed for different activities (**scalable**)
  - ◆ ... sometimes over large distances and with non-fault tolerant hardware (**robust**)
  - ◆ Maintain knowledge of what we are doing and what we did (**monitoring and bookkeeping**)
  - ◆ Use our limited resources effectively both at home and away (**efficient**)
  - ◆ Don't want to know the details [where files sit, where jobs run] (**transparent**)
  - ◆ Find data easily (**query tools**)
- 

## Solution...

- ◆ A data handling and job management system
- ◆ **SAMGrid**

# *What can SAMGrid do?*

---

- ◆ **SAMGrid manages file storage (replica catalogs)**
  - ❖ Data files are stored in tape systems at Fermilab and elsewhere. Files can be cached around the world for fast access
- ◆ **SAMGrid manages file delivery**
  - ❖ Users at Fermilab and remote sites retrieve files out of file storage. SAMGrid can handle caching for efficiency
  - ❖ You don't care about file locations
- ◆ **SAMGrid manages file metadata cataloging**
  - ❖ SAMGrid DB holds metadata for each file. You don't need to know the file names to get data
- ◆ **SAMGrid manages analysis bookkeeping**
  - ❖ SAMGrid remembers what files you ran over, what files you processed successfully, what applications you ran, when you ran them and where
- ◆ **SAMGrid manages jobs**
  - ❖ Choose execution site, deliver job, executables, and its needed data, store output

# *Deployment status*

---

- ◆ DØ: For DH, SAM is used exclusively
  - ❖ Reprocessing effort using JIM (job management part of SAMGrid)
  - ❖ Writing interoperability with LCG code
  - ❖ Starting to deploy v6 at FNAL
  
- ◆ CDF: Deploying SAM to be main DH
  - ❖ Integration and load testing with v6 on CAF
  - ❖ Remote sites have been using v5; converting to v6
  - ❖ Prototype SAM Farm is functioning
  - ❖ Deploying several SAMGrid [JIM] job execution sites for MC (Valeria)
  - ❖ See Krzysztof's talk
  
- ◆ MINOS: In the initial stages of deployment

# Goals

---

- ◆ Near term
  - ❖ Serve the experiment customers
- ◆ Medium term
  - ❖ Transition to maintenance mode for experiment customers; collaborate with Grid stakeholders & concentrate on interoperability
- ◆ Long term
  - ❖ Contribute components to Grid systems
- ◆ *We are working on many of these aspects **now***

# v6 SAM

---

- ◆ v6 SAM
  - ❖ Make full use of new DB Schema
  - ❖ Rewrite of DB Server (middleware) - more robust, performant, and maintainable
  - ❖ Improvements to SAM client, station, DDE
- ◆ CDF is deploying
- ◆ DØ starting to deploy
  - ❖ Need to work transition into experiment schedule



# *Near Term Goals*

---

- ◆ Big stuff
  - ❖ Deploy SAM DH at CDF
    - Gerry Gugliemo is “consultant”
  - ❖ Deploy SAMGrid at some CDF sites for MC
  - ❖ Deploy v6 at DØ
  - ❖ Run SAMGrid for DØ remote reprocessing
  - ❖ Deploy SAM at MINOS
  
- ◆ Additional ongoing projects
  - ❖ Issue tracker
  - ❖ Improve packaging/configuration/installation
  - ❖ SQL Builder
  - ❖ Querying distributed databases (SBIR II)
  - ❖ Station code evaluation
  - ❖ SAMGrid/LCG interoperability

# *Near Term Goals*

---

## ◆ MIS server

### ❖ Next generation monitoring (not log files)

- SAMGrid applications send monitoring events to server
- Backend applications process events and store, display, or take action
- Extremely flexible system - tailored for SAMGrid

### ❖ Information cache

- Improve efficiency and performance by caching DB queries near the SAM DH stations

### ❖ Status

- Monitoring server 98% complete
- SAMGrid application instrumentation 90% complete
- SamTV transition to SamHDTV 50% complete
- Information cache ongoing

# *Near Term Goals*

---

- ◆ Future projects to start soon
  - ❖ RCP to SAM Parameters (CDF Request)
  - ❖ Process bookkeeping (CDF/DØ Request)
  - ❖ MC/Farm request system (CDF/DØ Request)
  - ❖ Test harness improvements
  - ❖ SAMGrid Job Brokering
  - ❖ SAMGrid Job Handling for Analysis
  - ❖ MISWEB enhancements (CDF Request)
  - ❖ SAM SRM interface
  - ❖ Improve cache groups and quotas (Potential Request)
  - ❖ Document what we've got in a journal
  - ❖ New requests from experiments
  
- ◆ Maintenance and operations support ongoing
  - ❖ Experts support SAM shifters and react quickly to deployment issues

# *Effort*

---

- ◆ Somewhat in flux, a snapshot...

- ❖ 6.0 FTEs development (SAMGrid team)

- ❖ +2 possible FTEs from new hires

- ❖ 3 FTEs non-FNAL funds

- ❖ 7.5 FTEs operations and deployment

## *Medium Term Goals*

---

- ◆ After SAMGrid matures at experiments, new feature requests should diminish
- ◆ Feature freeze SAMGrid for experiments
- ◆ Explore deeper interoperability with Grid systems; collaborate
  - ❖ More integrated VO system (authorization, roles)
  - ❖ Jobs using SAM DH can run on other Grid systems (already started for LCG)
  - ❖ Jobs running on other Grid systems but not explicitly using SAM DH have access to SAM services via “standard” Grid tools

## *Long term goals*

---

- ◆ Can SAMGrid contribute components of functionality to Grid systems?
  - ❖ Meta-data catalog
  - ❖ Caching and file transfer routing
  - ❖ Bookkeeping (request systems)
  - ❖ Job brokering and monitoring
  
- ◆ Would require breaking SAMGrid into pieces
  - ❖ Much work for DH components, since is a monolithic system -- but worth exploring
  - ❖ Export SAMGrid as a Fermilab product

# Summary

---

- ◆ Run II experiments and MINOS will all soon rely on SAMGrid for their data handling needs
  - ❖ Working well at DØ, soon to be deployed at CDF and MINOS
  - ❖ Support of the experiments must be primary goal for the near - medium term
- ◆ SAMGrid is a unique product
  - ❖ A Grid system doing real physics work today; not a prototype
  - ❖ Exporting our knowledge and perhaps code to other Grid systems is a longer term goal